



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2020

---

## **The performance effects of wise and unwise managerial dismissals**

Flepp, Raphael ; Franck, Egon

**Abstract:** The decision to dismiss a coach is challenging because poor performance tends to coincide with both bad luck and low coaching ability. We differentiate between dismissals following actual poor performance on the pitch (wise dismissals) and dismissals following seemingly poor performance due to bad luck (unwise dismissals). To categorize dismissals, we use “expected goals,” which are less vulnerable to random variation in match outcomes. Using data from European football, we find that wise dismissals increase subsequent performance compared to a control group of nondismissals with similarly poor performance on the pitch. However, unwise dismissals do not improve subsequent performance compared to a control group with similar strings of bad luck.

DOI: <https://doi.org/10.1111/ecin.12924>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-188874>

Journal Article

Accepted Version

Originally published at:

Flepp, Raphael; Franck, Egon (2020). The performance effects of wise and unwise managerial dismissals. *Economic Inquiry*, 59(1):186-198.

DOI: <https://doi.org/10.1111/ecin.12924>

# The performance effects of wise and unwise managerial dismissals

Raphael Flepp\*

Egon Franck\*\*

## Abstract

While dismissing a coach is one of the key decisions made by football club boards, this decision is challenging because poor performance tends to coincide with both bad luck and low coaching ability. Our paper is the first to differentiate between coach dismissals following actual poor performance on the pitch (wise dismissals) and coach dismissals following seemingly poor performance due to bad luck (unwise dismissals). To categorize dismissals, we use "expected goals", which are less vulnerable to random variation in match outcomes. Using match-level data from the big five European football leagues, we find that wise dismissals increase subsequent performance compared to a control group of non-dismissals with similarly poor performance on the pitch. However, unwise dismissals do not improve subsequent performance compared to a control group of non-dismissals with similar strings of bad luck. Our results have important implications for the design of future coach dismissal studies and for football clubs to reduce ineffective dismissal decisions.

JEL Classification: D81, J44, L83

This is the peer reviewed version of the following article: Flepp, R. and Franck, E. (2020). The performance effects of wise and unwise managerial dismissals. *Economic Inquiry*, which has been published in final form at <http://dx.doi.org/10.1111/ecin.12924>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

---

\*Corresponding author. Department of Business Administration, University of Zurich, Switzerland. E-mail: [raphael.flepp@business.uzh.ch](mailto:raphael.flepp@business.uzh.ch)

\*\*Department of Business Administration, University of Zurich, Switzerland. E-mail: [egon.franck@business.uzh.ch](mailto:egon.franck@business.uzh.ch)

# 1 Introduction

Understanding how managerial change affects the performance of organizations is of critical importance. One main area of previous research is the sports industry, where the focus is on the impact of head coach changes on team performance (Giambatista, Rowe, & Riaz, 2005). Coaches are ultimately responsible for team performance and thus hold official authority (Grusky, 1963). Based on the considerable influence of coaches on teams through strategic, operational and motivational channels, research suggests that coaches play crucial roles in determining team performance (e.g., Rowe, Cannella Jr, Rankin, & Gorman, 2005; Muehlheusser, Schneemann, Sliwka, & Wallmeier, 2018; Kattuman, Loch, & Kurchian, 2019).

However, most empirical studies across different sports conclude that there is no positive effect of involuntary coach changes on team performance (e.g., Gamson & Scotch, 1964; Audas, Goddard, & Rowe, 2006). Most prominently, studies using data from European football show that while performance improves after a within-season coach dismissal, performance also improves for a control group of counterfactual dismissals that did not occur (e.g., De Paola & Scoppa, 2012; van Ours & van Tuijl, 2016). Thus, even though within-season dismissals occur frequently and usually attract considerable media attention, they do not seem to cause improved team performance; rather, mean reversion drives subsequent performance improvements.

One potential explanation for this puzzling finding is that football club boards make dismissal decisions based on misperceptions. While dismissing a coach is one of the key decisions made by football club boards, this decision is challenging because poor performance tends to coincide with bad luck and low manager ability (Huson, Malatesta, & Parrino, 2004). Thus, boards may falsely infer manager ability from exogenous performance components, which could disguise any positive replacement effect. Indeed, Kaplan and Minton (2012) and Jenter and Kanaan (2015) show that even professional corporate

boards fail to filter out exogenous elements, and top managers are dismissed after poor firm performance caused by factors beyond their control.

These misperceptions are likely to be amplified for the boards of football clubs. Because football is a low-scoring sport, the random component in a single game is considerable. Anderson and Sally (2014, p. 40) even argue that “in fact [the role of chance] is about 50/50. Half of the goals you see, half of the results you experience, are down not to skill and ability but to random chance and luck.” Several studies have shown that decision makers underestimate the role of randomness in match outcomes and assign too much weight to the observed outcomes when they evaluate performance, a phenomenon that is commonly referred to as outcome bias (e.g., Lefgren, Platt, & Price, 2015; Gauriot & Page, 2019). Consequently, the dismissal decisions of football club boards are likely based on misperceptions when exogenous factors shape match outcomes (Brecht & Flepp, 2020). In such cases, club boards may systematically attribute bad luck to poor coach quality.

This paper is the first to account for boards’ misperceptions when dismissing coaches by differentiating between coach dismissals following actual poor performance on the pitch (*wise dismissals*) and coach dismissals following seemingly poor performance due to bad luck (*unwise dismissals*). If head coaches matter and replacing a bad coach with a more skilled one has an actual impact, after wise dismissals, we would expect team performance to improve compared to that of a control group of teams with similarly poor performance on the pitch but for which there was no dismissal. After unwise dismissals, however, we do not expect team performance to improve compared to that of a control group of non-dismissal teams that faced similar bad luck, as performance reverts to the mean in both groups.

We use match-level data from the English Premier League, the French Ligue 1, the German Bundesliga, the Italian Serie A and the Spanish La Liga for the five seasons from 2013/2014 to 2017/2018. During this sample period, we registered 143 involuntary in-season coach dismissals. To differentiate between wise and unwise dismissals, we compare

the ranking of teams in the official league table to the ranking of teams based on "expected goals" at the time of the dismissal. Following Brechot and Flepp (2020), we derive the expected goals (xG) for a match by summing up the estimated scoring probability for each shot.<sup>1</sup> For example, if a team took five shots in a match with estimated scoring probabilities of 0.05, 0.10, 0.10, 0.60, and 0.90, the number of expected goals equals 1.75. Because shots occur much more frequently than goals, xG is a performance evaluation measure that is less prone to randomness. Indeed, Brechot and Flepp (2020) show that the xG metric reflects true performance on the pitch more accurately than does the match outcome. Consequently, a ranking based on xG should reflect a team's playing quality on the pitch better than its official league ranking because the latter is fully subject to the random component of match outcomes.

In our main specification, we classify a dismissal as wise if the rank based on xG is equal to or worse than the rank in the official league table. In such situations, teams fall short of expectations due to poor performance on the pitch, and replacement of the coach appears justified. If the rank based on xG is better than the rank in the league table, we classify the dismissal as unwise because the team likely performed below expectations due to bad luck. Conversely, we classify all potential non-dismissals, i.e., all observations of teams without dismissals during a particular season, as wise if the rank based on xG is equal to or worse than the rank in the official league table, and as unwise otherwise.<sup>2</sup>

To account for biased club board decisions and mean reversion effects, we compare dismissals after poor performance on the pitch to non-dismissals after poor performance on the pitch (i.e., wise dismissals vs. unwise non-dismissals) and dismissals after bad luck to non-dismissals after bad luck (i.e., unwise dismissals vs. wise non-dismissals). Building upon the methods of van Ours and van Tuijl (2016), we match wise dismissals to unwise non-dismissals and unwise dismissals to wise non-dismissals based on the same team and a similar "cumulative surprise", i.e., the sum of the difference between the number of actual

---

<sup>1</sup> In Section 3.2, we explain the estimation of shot probabilities in detail.

<sup>2</sup> Our results are robust to alternative cutoffs for wise and unwise dismissals (see Section 4.2).

points and the expected points derived from betting odds. Thus, each control group performed similarly below expectations due to either poor performance on the pitch or a string of bad luck.

Using a team-season fixed effects regression model, we find that team performance significantly increases after wise dismissals but not after unwise non-dismissals. In particular, the F-test for equality of parameters is rejected, which implies that replacing a coach has a positive effect if the team performed below expectations due to poor playing on the pitch. By contrast, we find that team performance similarly increases after both unwise dismissals and wise non-dismissals. Thus, in situations where a team performed below expectations due to bad luck, team performance reverts to the mean regardless of whether the coach is dismissed. Furthermore, employing a matching method that omits the requirement that each dismissal must be matched to a non-dismissal of the same team leads to the same results.

Our paper makes several contributions to the literature on the effects of coach changes on team performance. First, we account for club boards' biased decision making and show that post-dismissal performance critically depends on the situation preceding the dismissal decision. Replacement of a coach is beneficial only if poor performance on the pitch was indeed the driver of below expectation performance. If, however, a replacement decision was triggered when a string of bad luck was the main cause of performing below expectations, the replacement of the coach has no effect on subsequent performance. While Brechot and Flepp (2020) discuss a few examples wherein teams were ranked very differently by the official league table than by expected goals, we systematically focus on subsequent performance effects of wise and unwise (non-)dismissals. Second, our results explain why many previous studies failed to find causal performance improvements after coach dismissals. Considering that a substantial fraction of club boards' dismissal decisions are unwise, the average post-performance effect of a coach dismissal is biased toward zero due to the effect of mean reversion. Thus, we advance the insights of van Ours and

van Tuijl (2016) by applying their methods in light of comparability but critically accounting for biased club boards' dismissal decisions. Finally, the disclosure of suboptimal coach dismissal decisions that trigger financial consequences without performance effects is complex. Because post-dismissal performance also improves after unwise dismissals, boards may view their decisions as justified even though the same result could have been achieved without replacing the coach. Thus, football clubs should complement their coach evaluation strategies with more analytical approaches to reduce ineffective decisions. Replacing a head coach is expensive, and the costs associated with unwise dismissals could be better invested in new players to improve the strength of the team.

The remainder of this paper is structured as follows: In Section 2, we review the literature and derive the hypotheses. In Section 3, we present our empirical methods. In Section 4, we present the results and in Section 5, we conclude.

## **2 Related literature and hypotheses**

The question of whether coaches matter for team performance has been addressed by several studies. In a comprehensive study, Berry and Fowler (2019) find that coaches matter in Major League Baseball, the National Basketball Association (NBA), the National Hockey League (NHL), and the National Football League (NFL), as well as in college football and college basketball. Specifically, their results imply that coaches explain approximately 20-30% of the variation in a team's performance, depending on the sport and the performance outcomes used. Furthermore, Muehlheusser et al. (2018) analyze the impact of football coaches in the German Bundesliga. The authors find that coaches have a significant impact on team performance and that there is considerable ability heterogeneity among coaches. Moreover, Muehlheusser et al. (2018) show that coaches have a significant effect on teams' playing styles, for example, by employing more offensive or defensive players.

Kattuman et al. (2019) present a conceptual framework of how a football coach can influence a team and its performance. As inputs, the coach selects the talent on the field and sets the game strategy. Furthermore, he or she can drive motivation through his or her leadership and feedback style. These inputs potentially enhance team processes regarding physical aspects of training sessions, such as shooting or fitness, and motivational aspects, such as team cohesion, communication or emergent team states. Consequently, the coach is responsible for team development and nurturing talent (Hughes, Hughes, Mellahi, & Guermat, 2010; Pieper, Nüesch, & Franck, 2014). The chosen tactics, e.g., the speed of play or the system of positions, and the fielded players, in turn, influence game behavior. Finally, the combination of inputs and team processes influences the affect and effort level in the game, which ultimately translates into better team performance (Kattuman et al., 2019).

Based on the considerable influence of the head coach on the team, research on coach changes in professional football conjectures that coaches play a crucial role in determining team performance (e.g., Koning, 2003; Bryson, Buraimo, & Simmons, 2018; De Paola & Scoppa, 2012). Consequently, if coaches matter and a team performs poorly, replacing an ineffective coach with an effective one should improve team performance (van Ours & van Tuijl, 2016). In this context, the literature has acknowledged that poor performance tends to coincide with bad luck and low manager ability (Huson et al., 2004). Thus, an improvement in subsequent performance could be due either to the ability of the new coach or to mean reversion. To account for this possibility, it is crucial to compare actual dismissals to a control group of non-dismissals with similarly poor prior performance (Giambatista et al., 2005).

De Paola and Scoppa (2012) employed nearest neighbor matching based on the ranking difference, the number of points obtained in the four most recent matches and the period in a particular season to find a counterfactual dismissal for each actual dismissal. Using 12 seasons from the Italian Serie A between 1997/98 and 2008/09, De Paola and Scoppa



(2012) find that improved team performance is due solely to mean reversion; thus, a coach change does not causally affect team performance. This finding is in line with those of several other studies using control groups of counterfactual dismissals (e.g., Baldock, Buelens, & Philippaerts, 2010; Heuer, Müller, Rubner, Hagemann, & Strauss, 2011; Ter Weel, 2011).

More recently, van Ours and van Tuijl (2016) formed a control group by matching a team to itself based on cumulative surprise, which measures how far below expectations the team performs and is calculated as the sum of the difference between the number of actual points and the expected points derived from betting odds. Using 36 dismissals and 33 counterfactual dismissals from the Dutch Eredivisie in the 14 seasons between 2000/01 and 2013/14, the authors find that forced coach replacements do not improve team performance. Besters, van Ours, and van Tuijl (2016) replicate the methods of van Ours and van Tuijl (2016) using 45 dismissals and 34 counterfactual dismissals from the English Premier League between 2000/01 and 2014/15. Similarly, they conclude that, on average, performance does not improve following a coach’s dismissal and that a successful coach change remains highly unpredictable. Overall, most empirical evidence from professional football shows that coach dismissals have no causal effect on subsequent team performance.<sup>3</sup>

However, the role of football club boards’ misperceptions when replacing a coach has been neglected in the literature. If the coach evaluations of club boards are partly driven by irrelevant cues such as unlucky losses (Arnulf, Mathisen, & Hærem, 2012), it is likely that club boards systematically attribute bad luck to poor coach quality when making dismissal decisions. Thus, to properly test whether a coach’s dismissal is beneficial for a football club, we differentiate between dismissals and non-dismissals that occur due to either poor performance on the pitch or simply bad luck. We refer to a dismissal as *wise*

---

<sup>3</sup> There exist, however, some exceptions. For example, Madum (2016) finds a positive performance effect but only for home games; Muehlheusser, Schneemann, and Sliwka (2016) find that dismissals enhance the performance of homogeneous teams but not of heterogeneous teams; and Bryson et al. (2018) find a positive performance effect after dismissals but no effect after voluntary quits.

if a team performed below expectations due to poor performance on the pitch and as *unwise* if a team performed below expectations due to bad luck. Conversely, we refer to a counterfactual non-dismissal as *wise* if a team performed below expectations due to bad luck and as *unwise* if a team performed below expectations due to poor performance on the pitch. Table 1 illustrates this theoretical decomposition of dismissals and non-dismissals.

**Table 1**  
Theoretical decomposition of dismissals and non-dismissals.

	dismissal	non-dismissal
wise	Below expectations due to poor performance on the pitch	Below expectations due to bad luck
unwise	Below expectations due to bad luck	Below expectations due to poor performance on the pitch

To gain insight from this decomposition, dismissals after poor performance on the pitch must be compared to non-dismissals after poor performance on the pitch, and dismissals after bad luck must be compared to non-dismissals after bad luck. We expect that a dismissal benefits a team if a coach whose team played poorly on the pitch is replaced. Thus, we formulate the first hypothesis as follows:

**H1.** *Wise dismissals improve performance relative to a control group with similarly poor performance on the pitch but no dismissal (i.e., unwise non-dismissals).*

We expect performance following an unwise dismissal to improve in a manner similar to that following a wise non-dismissal because of simple mean reversion. Thus, we hypothesize the following:

**H2.** *Unwise dismissals do not improve performance relative to a control group with similarly bad luck but no dismissal (i.e., wise non-dismissals).*

In the next section, we present our methods, outline how we classify dismissals as either wise or unwise and explain how we construct the control groups of (non-)dismissals to test our hypotheses.

## 3 Methods

### 3.1 Data

We employ data from the English Premier League, the French Ligue 1, the German Bundesliga, the Italian Serie A and the Spanish La Liga for the five seasons from 2013/2014 to 2017/2018. Our main data set contains 9,130 matches for which we collected the date, the teams, the final score, and the head coach of each team from [www.transfermarkt.com](http://www.transfermarkt.com) and [www.football-data.co.uk](http://www.football-data.co.uk). For every in-season coach change, we recorded whether it was a dismissal or a voluntary quit. Following Pieper et al. (2014), we categorize a coach change as a dismissal if the club's board unilaterally terminated the employment and categorize a coach change as voluntary quit if the coach unilaterally resigned or if the termination was with mutual consent.<sup>4</sup> As in van Ours and van Tuijl (2016), we ignore coach dismissals in the first four and the last four match weeks because we aim to estimate the difference in team performance before and after a dismissal. Further, following Besters et al. (2016), we consider only the first coach dismissal in a season.

As a result, we registered 143 coach dismissals during our sample period. The data were then complemented by the betting odds for each match and the final rank of each club in the previous season.<sup>5</sup> Additionally, we employ a second data set consisting of

---

<sup>4</sup> Our results do not change if we reclassify the terminations explained as "by mutual consent" as dismissals instead of voluntary quits.

<sup>5</sup> We downloaded the betting odds from [www.football-data.co.uk](http://www.football-data.co.uk). As in van Ours and van Tuijl (2016), our bookmaker data are mainly from William Hill (99.96%) and supplemented by Ladbrokes (0.04%). The final rank in the previous season for promoted teams is set to 20 (18 for the German Bundesliga).

all 214,194 shots taken during the 9,130 matches included in our sample to estimate the expected goals performance metric.<sup>6</sup>

### 3.2 Categorization of (non-)dismissals using expected goals

Football is a low-scoring game, and a team might unexpectedly lose or tie a game despite playing well. To differentiate situations in which teams performed below expectations due to poor performance on the pitch from situations in which teams performed below expectations due to a string of bad luck, we need a measure of performance that is less prone to random variation.

We draw on the concept of expected goals (xG) based on quantified scoring chances. Scoring chances are the second-to-last step in the goal production process, and all teams try to create valuable chances to score goals. Following Brechot and Flepp (2020), we employ shots as scoring chances and estimate their scoring probability. In our sample, we observe a total of 214,194 shots that translated into 24,082 goals scored.<sup>7</sup> Thus, the average number of shots in a match is approximately 23.5, whereas the average number of goals is approximately 2.7. Because shots occur much more often than goals, the expected goal metric is less vulnerable to the randomness associated with actual goals and match results.

To test the superior informativeness of the xG metric, Brechot and Flepp (2020) examine how well past match outcomes and past expected goal performance explain the future success of teams. In particular, they regress the future number of points won separately on the past number of points won and on the past expected goal difference and compare the resulting  $R^2$  values. Using different combinations of the number of previous matches and the number of following matches, Brechot and Flepp (2020) show that  $R^2$  is higher when the xG metric is used. For example, when the ten previous matches are used to

---

<sup>6</sup> The shot data are provided by Nielsen, a measurement and data analytics company. The data for the seasons from 2013/2014 to 2015/2016 are the same as those used in Brechot and Flepp (2020).

<sup>7</sup> Own goals are excluded because they are unintended.

predict the following ten matches, the past number of points won explains 25.3% of the variation in future points, whereas the past expected goal difference explains 32%. Thus, the xG metric increases  $R^2$  by 6.7 percentage points (or 26.5%). Overall, Brechot and Flepp (2020) conclude that the xG metric, even though it is based on a simple model used to estimate the scoring probability of a shot, reflects true performance on the pitch more accurately than does the match outcome.

We estimate the scoring probability of an individual shot by running a logistic regression model where the binary response variable indicates whether or not the shot translated into a goal. As explanatory variables, we include the distance to the goal, the angle to the goal, the rule setting (i.e., open play, free kick or penalty kick), and the body part used for each shot. As Brechot and Flepp (2020) emphasize, the estimated scoring probabilities are likely to be biased if goal scoring or goal conceding qualities of teams are not fully captured by the model. To mitigate this bias, we include team-season and opposing team-season fixed effects in addition to the model developed by Brechot and Flepp (2020). These fixed effects account for time-constant heterogeneity in team quality characteristics, such as offensive and defensive skills within a particular season. In Appendix A, we provide more detailed variable descriptions, present descriptive statistics, outline the logistic regression model, and show the estimated model parameters. Based on these parameters, we predict the scoring probabilities and aggregate them for each team within a match to derive the number of expected goals scored per match. For example, if a team took three shots in a match with estimated scoring probabilities of 0.70, 0.15, and 0.05, the expected goals scored are equal to 0.90.

The expected goal metric allows us to determine which team on the pitch is better at creating valuable scoring chances. Instead of awarding the number of points based on the actual goals scored in a match, i.e., three for a win, one for a draw and zero for a loss, we award points based on expected goals. A team wins the match and is awarded three points based on expected goals if the expected goal difference is larger than 0.5. If the

expected goal difference is between -0.5 and 0.5, the match ends in a draw, and each team receives one point. Otherwise, the team loses the match based on expected goals, and no points are awarded.

Table 2 displays descriptive statistics and correlations for outcome-based performance measures and expected goal-based performance measures for the 5,326 team-matches played by teams for which one of the 143 dismissals occurred in a particular season.<sup>8</sup> On average, the number of goals scored and the goal difference are very similar to the number of expected goals scored and the expected goals difference. However, the standard deviations of the latter two are considerably lower. The average number of points won is 1.11, whereas the number of points based on expected goals is 1.00. The correlation between these two measures is 0.45, which seems reasonable given the premise that a considerable fraction of match outcomes and, thus, the number of actual points are driven by luck. Using the number of points based on expected goals should, at least partly, filter out the random component of lucky and unlucky match outcomes.

**Table 2**  
Descriptive statistics and correlations.

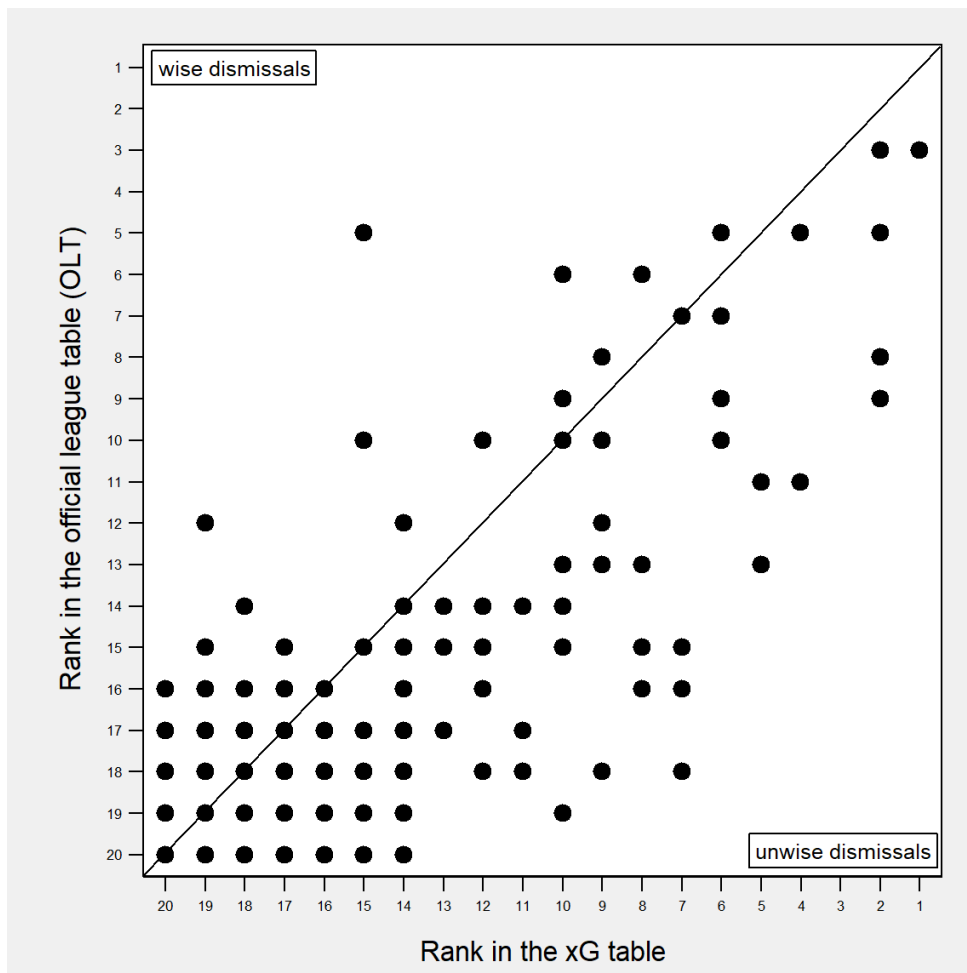
		Mean	Std. dev.	Min.	Max.	1	2	3	4	5	6
1	<i>Goals scored</i>	1.11	1.11	0	10						
2	<i>Goal diff</i>	-0.42	1.81	-8	8	0.67					
3	<i>Win</i>	0.29	0.45	0	1	0.61	0.74				
4	<i>Points</i>	1.11	1.26	0	3	0.63	0.84	0.95			
5	<i>xG scored</i>	1.11	0.74	0	7.77	0.56	0.42	0.34	0.36		
6	<i>xG diff</i>	-0.41	1.30	-6.94	7.11	0.39	0.60	0.41	0.47	0.69	
7	<i>Points based on xG</i>	1.00	1.15	0	3	0.38	0.50	0.41	0.45	0.67	0.81

Notes: N = 5,326 team-matches played by teams where one of the 143 dismissals occurred in a particular season.

To identify situations where match outcomes misrepresent performance on the pitch, we rank teams by points assigned based on expected goals and construct an xG league table. The rank in the xG table should reflect a team's performance on the pitch more accurately than the rank in the official league table (OLT) because the OLT draws solely on actual match outcomes, where bad luck directly translates into fewer points and a lower rank. For example, a team could play well on the pitch and win a match in terms of

<sup>8</sup> A team-match refers to a match from the perspective of the focal team.

expected goals because that team created more valuable scoring chances than its opponent. However, that team might actually lose the match if scoring chances did not translate into goals. In such situations, the xG table rank should be better than the OLT rank. In other situations where a team played poorly, the xG table rank should be equal to or worse than the OLT rank. Building upon Brechot and Flepp (2020), we plot the OLT rank against the xG table rank, which is useful for visualizing situations in which randomness likely played a considerable role in match outcomes. Figure 1 shows the rank combinations of all 143 dismissals at the time the dismissal occurred.<sup>9</sup> Notably, the correlation between the two rankings is approximately 0.75.



**Figure 1**  
Rank in the official league table vs. rank in the xG table at the time of coach dismissal. The identity line marks the cutoff for wise and unwise dismissals.

<sup>9</sup> Some dots represent multiple dismissals with the same rank combination.

To categorize a dismissal as wise or unwise, we use the difference between the xG table rank and the OLT rank. In our baseline analysis, we employ a rank difference of zero as the cutoff value.<sup>10</sup> Thus, we categorize a dismissal as wise if the rank based on xG is equal to or worse than the rank based on the OLT. If the rank based on xG is better than the rank based on the OLT, we categorize a dismissal as unwise. Such teams performed below expectations due to bad luck because their xG-based ranking indicates higher quality play on the pitch. The identity line in Figure 1 marks the cutoff for wise and unwise dismissals. Dismissals above or on the line are categorized as wise, whereas dismissals below the line are classified as unwise.

Conversely, we categorize all potential non-dismissal observations, i.e., all observations that are not associated with a team-season wherein an actual dismissal occurred, as wise if the ranking based on xG is equal to or better than the ranking based on the OLT, and as unwise otherwise. This categorization of wise and unwise non-dismissals is needed to construct an appropriate control group for wise and unwise actual dismissals.

### 3.3 Treatment and control group matching

Because dismissals are not exogenous, we aim to construct a control group of teams without a dismissal but otherwise identical characteristics for both wise and unwise dismissals to infer their causal effect on team performance. We build on the methods of van Ours and van Tuijl (2016) to match wise and unwise dismissals (treatment groups) to wise and unwise non-dismissals (control groups). In particular, we match wise dismissals to unwise non-dismissals and unwise dismissals to wise non-dismissals based on the same team and the closest cumulative surprise (CS) using the nearest neighbor approach. The CS measures deviations from expectations and is defined as the sum of the differences between

---

<sup>10</sup> In the sensitivity analysis in Section 4.2, we test alternative rank difference cutoffs of -1, -2, and -3.



the actual number of points won in a match and the expected number of points based on betting odds.<sup>11</sup>

To qualify for inclusion in a control group, a non-dismissal must stem from the same team but another season without a coach change, and the CS values must not differ by more than 0.5.<sup>12</sup> Matching based on the same team accounts for unobserved heterogeneity among teams because some teams might be more likely than other teams to dismiss a coach under similar circumstances (van Ours & van Tuijl, 2016). Further, this matching controls for time-constant seasonal aspirations of teams, such as qualifying for the UEFA Champions League or avoiding relegation. Moreover, matching counterfactual situations based on CS ensures that the control team performed similarly below expectations without dismissing the coach afterwards. Following this matching procedure, we were able to match 41 dismissals to non-dismissals. Figure 2a displays the matched dismissals and non-dismissals based on this restricted team matching procedure using the identity line as the cutoff for wise and unwise (non-)dismissals.

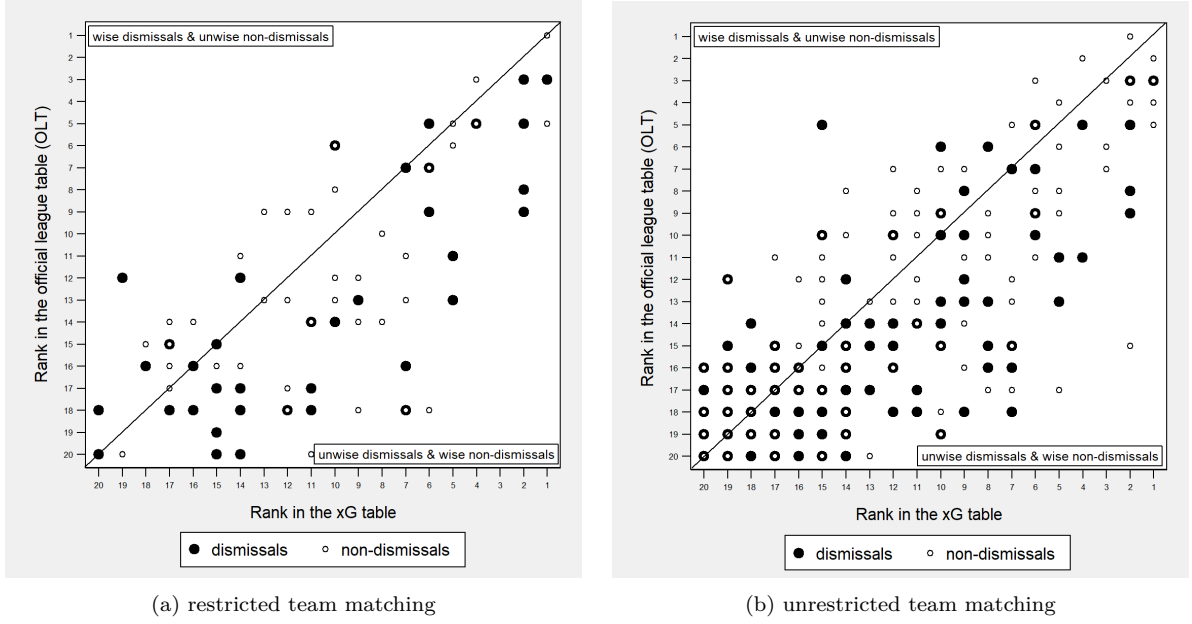
As an alternative matching method, we follow De Paola and Scoppa (2012) and Bryson et al. (2018) and omit the restriction that each dismissal has to be matched to a non-dismissal of the same team.<sup>13</sup> Thus, for every (un)wise dismissal, we match a corresponding (un)wise non-dismissal using only the CS. Employing this procedure, we are able to match all 143 dismissals to non-dismissals. Figure 2b shows the matched dismissals and non-dismissals based on this unrestricted team matching procedure, again using the identity line as the cutoff for wise and unwise (non-)dismissals.

---

<sup>11</sup> The expected number of points is calculated as  $[(\text{probability of win}) \cdot 3] + [(\text{probability of draw}) \cdot 1]$ . The probabilities are calculated as the reciprocal value of the corresponding betting odds and are adjusted for the bookmaker's over-round. For more detailed information, see van Ours and van Tuijl (2016). Note that the expected number of points differs fundamentally from the number of points based on expected goals. The latter is based on actual performance on the pitch *after* the game, whereas the former is based on expectations *prior to* the game.

<sup>12</sup> As for actual dismissals, we ignore non-dismissals within the first four and last four matches within a season. Alternative maximum CS differences of 0.25, 0.75 and 1.0 do not alter our main conclusions.

<sup>13</sup> Furthermore, the control group employed by Heuer et al. (2011) can consist of any team that displays a similar goal difference.



**Figure 2**

Matched dismissals and non-dismissals. Restricted team matching is based on the same team, whereas unrestricted team matching allows for matching across different teams. The identity line marks the cutoff for wise and unwise (non-)dismissals.

### 3.4 Econometric specification

Building upon the approach of van Ours and van Tuijl (2016), we estimate the regression model outlined in Equation 1:

$$\begin{aligned}
 \gamma_{ijk} = & \beta_1 \text{Wise dismissal}_{ijk} + \beta_2 \text{Unwise non-dismissal}_{ijk} \\
 & + \beta_3 \text{Unwise dismissal}_{ijk} + \beta_4 \text{Wise non-dismissal}_{ijk} \\
 & + \gamma_1 \text{Home}_{ijk} + \gamma_2 \text{Rank opponent}_{ijk} + \eta_{jk} + \epsilon_{ijk}
 \end{aligned} \tag{1}$$

where  $i$  denotes the match,  $j$  indicates the team and  $k$  refers to the season. For the dependent variable  $\gamma_{ijk}$ , we employ the same performance measures as in van Ours and van Tuijl (2016), namely, the number of points obtained in a match, a dummy variable indicating whether a match was won (*win*) and the goal difference in a match (*goal diff*). The (non-)dismissal variables are dummy variables that indicate whether there was a (non-)dismissal of the relevant type earlier in the season. We are mainly interested in the difference between the coefficients  $\beta_1$  and  $\beta_2$  and between the coefficients  $\beta_3$  and  $\beta_4$ .

According to our hypotheses, we expect  $\beta_1$  to be significantly larger than  $\beta_2$  (H1) and  $\beta_3$  to be equal to  $\beta_4$  (H2).

To account for unobserved differences in team quality within a particular season, we include team-season fixed effects  $\eta_{jk}$ . Furthermore, we control for home field advantage by including the dummy variable *Home* and proxy the strength of the opponent by controlling for the final rank of the opponent in the previous season (*Rank opponent*).

## 4 Results

### 4.1 Main results

Table 3 presents our main results for the four (non-)dismissal groups and the two matching procedures.<sup>14</sup> For the restricted team matching procedure, the number of observations is 2,854, including 1,522 matches from teams with a dismissal and 1,332 matches from teams with a non-dismissal.<sup>15</sup> Column (1) shows that after a wise dismissal, the number of points per match is 0.23 higher than before the dismissal. By contrast, the effect for the relevant control group of unwise non-dismissals is -0.10 points per match. The F-test for equality shows that these parameters are significantly different ( $F = 7.30^{***}$ ). The findings for *Win* and *Goal diff*, the performance measures in Columns (2) and (3), respectively, are similar, as the F-tests are again rejected. Thus, wise dismissals significantly improve performance relative to unwise non-dismissals.

With regard to unwise dismissals, Column (1) of Table 3 shows that the number of points per match significantly improves by 0.41. However, wise non-dismissals similarly improve team performance by 0.43 points per match. Indeed, the F-test for equality of these parameters cannot be rejected ( $F = 0.04$ ). This result suggests that the improvement

<sup>14</sup> We also replicated the approach of van Ours and van Tuijl (2016) by pooling all dismissals into one treatment group and all non-dismissals into one control group. Like van Ours and van Tuijl (2016), we fail to find any significant performance difference between the treatment group and the control group for both the restricted and the unrestricted matching procedures. Thus, the results of van Ours and van Tuijl (2016) are nested within our approach.

<sup>15</sup> Dismissal observations consist of 9×34 team-matches from the German Bundesliga and 32×38 team-matches from the other leagues. Non-dismissal observations consist of 9×34 team-matches from the German Bundesliga and 27×38 team-matches from the other leagues. Five of the 36 non-dismissal team-seasons serve as the control group for two dismissals.

**Table 3**  
Main results.

	restricted team matching			unrestricted team matching		
	<i>Points</i> (1)	<i>Win</i> (2)	<i>Goal diff</i> (3)	<i>Points</i> (4)	<i>Win</i> (5)	<i>Goal diff</i> (6)
<i>Wise dismissal</i>	0.23*** (0.080)	0.10*** (0.030)	0.18 (0.146)	0.23*** (0.042)	0.08*** (0.015)	0.26*** (0.065)
<i>Unwise non-dismissal</i>	-0.10 (0.095)	-0.02 (0.038)	-0.17 (0.111)	0.05 (0.052)	0.02 (0.021)	-0.09 (0.076)
<i>Unwise dismissal</i>	0.41*** (0.055)	0.14*** (0.026)	0.44*** (0.089)	0.44*** (0.040)	0.15*** (0.016)	0.46*** (0.058)
<i>Wise non-dismissal</i>	0.43*** (0.078)	0.15*** (0.026)	0.57*** (0.109)	0.36*** (0.053)	0.13*** (0.019)	0.45*** (0.080)
<i>Home</i>	0.42*** (0.045)	0.15*** (0.015)	0.68*** (0.060)	0.46*** (0.023)	0.14*** (0.008)	0.72*** (0.033)
<i>Rank opponent</i>	0.05*** (0.004)	0.016*** (0.002)	0.08*** (0.006)	0.05*** (0.002)	0.014*** (0.001)	0.08*** (0.003)
<i>Team-season fixed effects</i>	Yes	Yes	Yes	Yes	Yes	Yes
$R^2$ overall	9.06%	7.77%	11.22%	8.54%	6.56%	11.24%
N	2,854	2,854	2,854	8,280	8,280	8,280
F-test $\beta_1 = \beta_2$	7.30***	5.69**	3.49*	7.27***	5.69**	12.37***
F-test $\beta_3 = \beta_4$	0.04	0.04	0.85	1.67	0.94	0.01

Notes: The table reports the coefficients estimated using team-season fixed effects regression models. The cutoff for wise and unwise (non-)dismissals is at equal ranks in the xG table and in the OLT. Columns (1) to (3) show the results for restricted team matching, and Columns (4) to (6) show the results for unrestricted team matching. Standard errors are corrected for heteroskedasticity and reported in parentheses. In all models, \*, \*\*, and \*\*\* denote significance at the 10%, 5% and 1% levels, respectively.

in points per match after unwise dismissals would also have occurred if the coach had not been dismissed. The F-tests in Columns (2) and (3) confirm the finding that unwise dismissals do not improve performance relative to the relevant control group of wise non-dismissals.

For the unrestricted team matching procedure, in which all 143 actual dismissals are matched to a non-dismissal, the number of observations is 8,280, consisting of 5,326 matches from dismissal teams and 2,954 matches from non-dismissal teams.<sup>16</sup> Columns (4) to (6) show that the results remain similar. Specifically, the F-tests for equality show that the performance effect of wise dismissals is significantly larger than the performance effect of wise non-dismissals, but that there is no performance gain from unwise dismissals over wise non-dismissals.

<sup>16</sup> Again, several non-dismissal team-seasons serve as controls for multiple actual dismissals.

Note that unwise non-dismissals do not have a positive effect on subsequent performance in any of the specifications in Table 3. This result is expected if these teams indeed performed poorly on the pitch in the absence of bad luck and the coach was not replaced. Thus, the categorization into unwise non-dismissals using expected goals appears to successfully filter out the random component that usually leads to mean reversion in performance.

Overall, our main results support Hypothesis 1 that wise dismissals improve performance relative to an unwise non-dismissal control group with similarly poor performance on the pitch. Moreover, the positive performance effects observed for unwise dismissals appear to be driven solely by mean reversion, thus supporting Hypothesis 2.

## 4.2 Sensitivity analysis

The categorization into wise and unwise dismissals is crucial in our analysis. Even though the cutoff at equal OLT and xG table ranks is straightforward, several sensitivity tests are required to evaluate the robustness of our results. In particular, we do not know the details of the circumstances that triggered each decision to dismiss a coach. For example, conflicts with key players would be expected to translate into lower performance on the pitch in future matches. Thus, we might have labeled some dismissals as unwise even if the decision to dismiss the coach was justified.

We address this concern by testing alternative rank differences in favor of wise (non-)dismissals. First, we categorize a dismissal as wise even if the ranking based on xG is one spot higher than the OLT rank: only if the ranking based on xG is more than one spot higher do we categorize a dismissal as unwise. Analogously, we categorize a non-dismissal as unwise if the ranking based on xG is more than one spot lower than the ranking based on the OLT, and we categorize it as wise otherwise. Column (1) of Table 4 shows the results for the restricted team matching procedure using points as the dependent variable. Again, the F-tests reveal that wise dismissals significantly increase the number of points

obtained in a match compared to unwise non-dismissals, whereas unwise dismissals do not improve performance compared to wise non-dismissals.

**Table 4**  
Sensitivity analysis.

	<i>Points</i>					
	restricted team matching			unrestricted team matching		
	-1/+1 (1)	-2/+2 (2)	-3/+3 (3)	-1/+1 (4)	-2/+2 (5)	-3/+3 (6)
<i>Wise dismissal</i>	0.16** (0.072)	0.18** (0.077)	0.15 (0.129)	0.23*** (0.036)	0.27*** (0.038)	0.27*** (0.039)
<i>Unwise non-dismissal</i>	-0.09 (0.105)	-0.07 (0.085)	-0.13 (0.129)	0.02 (0.050)	-0.06 (0.056)	-0.09 (0.075)
<i>Unwise dismissal</i>	0.47*** (0.055)	0.44*** (0.054)	0.48*** (0.055)	0.50*** (0.044)	0.49*** (0.049)	0.47*** (0.051)
<i>Wise non-dismissal</i>	0.38*** (0.084)	0.34*** (0.081)	0.38*** (0.085)	0.33*** (0.062)	0.34*** (0.075)	0.33*** (0.084)
<i>Home</i>	0.43*** (0.048)	0.44*** (0.053)	0.44*** (0.063)	0.48*** (0.024)	0.48*** (0.026)	0.46*** (0.028)
<i>Rank opponent</i>	0.05*** (0.005)	0.05*** (0.005)	0.05*** (0.007)	0.04*** (0.002)	0.04*** (0.002)	0.04*** (0.003)
<i>Team-season fixed effects</i>	Yes	Yes	Yes	Yes	Yes	Yes
$R^2$ overall	8.66%	8.76%	8.55%	8.42%	8.29%	8.03%
N	2,444	2,012	1,328	7,870	7,160	5,812
F-test $\beta_1 = \beta_2$	3.93*	5.00**	2.37	11.73***	23.59***	18.65***
F-test $\beta_3 = \beta_4$	0.71	1.02	0.91	4.53**	2.65	1.82

Notes: The table reports the coefficients estimated using team-season fixed effects regression models based on varying rank difference cutoffs for wise and unwise (non-)dismissals. Columns (1) to (3) show the results for restricted team matching, and Columns (4) to (6) show the results for unrestricted team matching. Standard errors are corrected for heteroskedasticity and reported in parentheses. In all models, \*, \*\*, and \*\*\* denote significance at the 10%, 5% and 1% levels, respectively.

Next, we extend the rank difference in favor of wise (non-)dismissals to -2 and -3 for wise dismissals and to +2 and +3 for wise non-dismissals. Columns (2) and (3) of Table 4 show that the results remain robust to the -2/+2 cutoff. However, the difference between the performance effects of wise dismissals and unwise non-dismissals is not statistically significant for the -3/+3 cutoff (p-value = 0.13).<sup>17</sup>

Analogously, we assess the sensitivity of our results to varying cutoffs for the unrestricted team matching procedure. As shown in Columns (4) to (6) of Table 4, wise dismissals improve performance relative to unwise non-dismissals in all specifications, while

<sup>17</sup> Because Hypothesis 1 states that the performance effect of wise dismissals is larger than the performance effect of unwise non-dismissals, one could argue that a one-tailed F-test might also be feasible. In this case, the difference is significant at the 10% level.

unwise dismissals do not improve performance relative to wise non-dismissals in all but one specification.<sup>18</sup> Overall, the sensitivity analysis shows that our main findings broadly hold when using different cutoffs to categorize wise and unwise (non-)dismissals.<sup>19</sup>

## 5 Conclusion

This paper shows that the effect of head coach dismissals on performance critically depends on whether football club boards falsely attribute bad luck to low coach ability or whether performance was indeed poor when deciding to replace the coach. We find that dismissals after poor performance on the pitch increase subsequent team performance compared to the control group, while dismissals after bad luck do not.

Our study has important implications for the design of future studies that investigate the relation between coach dismissal and subsequent team performance. Depending on the importance of the luck and board misperception components in a certain sport, the average coach effect might be biased towards zero. Thus, without accounting for the misperceptions of boards in their coach dismissal decisions, the results might be misleading. Moreover, the findings have direct implications for football clubs. On the one hand, dismissing a head coach is expensive. First, the dismissed coaches receive indemnity payments for the instantaneous cancellation of the contract (Muehlheusser et al., 2016). Second, the dismissal of a head coach usually involves the departure of his backroom support staff. For example, assistant coach Giuseppe Carillo also departed US Palermo in

---

<sup>18</sup> Additionally, we run the complete sensitivity analysis with *Win* and *Goal diff* as dependent performance variables. The results remain unchanged except for the models using the goal difference in the restricted team matching procedure. While the coefficients of wise dismissals are still larger than the coefficients of unwise non-dismissals, the equality of the coefficients cannot be rejected.

<sup>19</sup> Furthermore, one might worry that the departure of assistant coaches disrupts the coaching staff and leads to declining performance even though the head coach remains in their post. In our analysis, such a mechanism could systematically lower the performance of teams in the unwise non-dismissal control group and explain why teams perform significantly better after wise dismissals than after unwise non-dismissals. To address this concern, we checked whether the assistant coaches of teams in the matched unwise non-dismissal control group left while the head coach remained. In all but one case, the assistant coaches also remained in their posts. The only exception is Steve Agnew who left Hull City during the 2014/15 season to become assistant coach at FC Middlesbrough. Our results remain robust to the exclusion of Hull City's 2014/15 season; thus, we are confident that the results are not driven by the departure of assistant coaches. Additionally, Peeters, Szymanski, and Terviö (2017) provide empirical evidence that clubs tend to rehire experienced managers even though a considerable fraction of those managers are of lower ability than novice head coaches. This preference to rehire experienced head coaches further complicates the promotion of assistant coaches to head coach positions.

the 2015/16 season and Udinese Calcio in the 2016/17 season when head coach Giuseppe Iachini was dismissed. Finally, there are recruiting costs associated with finding a new coach. Thus, instead of spending this money on unwise dismissals, it could be invested in the recruitment of better players. On the other hand, as wise dismissals do improve team performance, the money associated with such dismissals is likely to be well spent. Consequently, it is of critical importance for club boards to correctly assess the drivers that led to performance below expectations when considering the dismissal of a coach. In this regard, more analytical approaches, such as the use of expected goals, could help to reduce the number of ineffective dismissal decisions. Furthermore, football clubs could employ analytical approaches to justify their decisions to fans and the media, which might reduce the pressure on boards to dismiss coaches unnecessarily.

Certainly, our study has limitations. Most importantly, it is likely that the ranking based on expected goals is not perfectly accurate. While we include team-season and opposing team-season fixed effects to approximate team strength in our expected goals model, we cannot rule out that other factors not included in the model, such as defensive pressure or individual player skills, bias our xG estimates. However, as we compare the same teams to each other in the restricted team matching method, this bias should be present in both the dismissal groups and in the non-dismissal groups. Nevertheless, future studies should investigate more sophisticated or further analytical approaches to filter out the random component of team performance more rigorously.



# A Appendix

Table A.1 defines the variables used and shows the descriptive statistics for the 214,194 shots taken in the 9,130 matches in our sample.

**Table A.1**  
Descriptive statistics for shots.

Variable	Description	Mean	Std. dev.	Min.	Max.
<i>Goal</i>	Shot resulting in a goal (0/1)	0.112	0.316	0	1
Location					
<i>Distance</i>	Distance between the shot location and the midpoint between the two goalposts measured in meters	18.60	7.44	0.60	91.10
<i>Angle</i>	Angle between the shot location and the two goalpost measured in degrees	22.95	12.74	0.10	173.60
Rule setting					
<i>Open play</i>	Shot is taken during open play (0/1)	0.938	0.241	0	1
<i>Free kick</i>	Shot is taken as a free kick (0/1)	0.050	0.217	0	1
<i>Penalty kick</i>	Shot is taken as a penalty kick (0/1)	0.012	0.110	0	1
Body part					
<i>Foot</i>	Shot is taken with the foot (0/1)	0.842	0.365	0	1
<i>Header</i>	Shot is taken with the head or another body part (0/1)	0.158	0.365	0	1

**Notes:** The number of observation is 214,194.

Building on Brechot and Flepp (2020), we employ a logistic regression analysis and estimate the model as outlined in Equation A.1:

$$\begin{aligned}
 \text{Ln} \left[ \frac{P(\text{Goal}_{ijkl} = 1)}{P(\text{Goal}_{ijkl} = 0)} \right] = & \beta_1 \text{Distance}_{ijkl} + \beta_2 \text{Angle}_{ijkl} + \beta_3 \text{Free kick}_{ijkl} \\
 & + \beta_4 \text{Penalty kick}_{ijkl} + \beta_5 \text{Header}_{ijkl} + \theta_{jl} + \tau_{kl} + \varepsilon_{ijkl}
 \end{aligned} \tag{A.1}$$

where  $i$  denotes the shot,  $j$  and  $k$  indicate the team and the opposing team, respectively, and  $l$  refers to the season.  $\theta_{jl}$  and  $\tau_{kl}$  refer to team-season and opposing team-season fixed effects, respectively.

Table A.2 shows the parameter estimates for Equation A.1.

**Table A.2**

Estimation results from logistic regression.

	<i>Goal</i> (0/1)
<i>Distance</i>	-0.128*** (0.003)
<i>Angle</i>	0.032*** (0.001)
<i>Free kick</i>	1.014*** (0.046)
<i>Penalty kick</i>	2.203*** (0.047)
<i>Header</i>	-1.058*** (0.021)
<i>Team-season fixed effects</i>	Yes
<i>Opposing team-season fixed effects</i>	Yes
Pseudo $R^2$	0.204
N	214,194

**Notes:** Robust standard errors are given in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

## References

- Anderson, C., & Sally, D. (2014). *The numbers game: why everything you know about football is wrong*. Penguin UK.
- Arnulf, J. K., Mathisen, J. E., & Hærem, T. (2012). Heroic leadership illusions in football teams: Rationality, decision making and noise-signal ratio in the firing of football managers. *Leadership*, 8(2), 169–185.
- Audas, R., Goddard, J., & Rowe, W. G. (2006). Modelling employment durations of NHL head coaches: turnover and post-succession performance. *Managerial and Decision Economics*, 27(4), 293–306.
- Baldock, A.-L., Buelens, M., & Philippaerts, R. (2010). Short-term effects of midseason coach turnover on team performance in soccer. *Research Quarterly for Exercise and Sport*, 81(3), 379–383.
- Berry, C. R., & Fowler, A. (2019, March). How much do coaches matter? In *Sports Analytic Conference, MIT Sloan, Boston, MA*.
- Besters, L. M., van Ours, J. C., & van Tuijl, M. A. (2016). Effectiveness of in-season manager changes in English Premier League football. *De Economist*, 164(3), 335–356.
- Brechot, M., & Flepp, R. (2020). Dealing with randomness in match outcomes: how to rethink performance evaluation in European club football using expected goals. *Journal of Sports Economics*, 21(4), 335–362.
- Bryson, A., Buraimo, B., & Simmons, R. (2018). Special ones? The effect of head coaches on football team performance. *Working Paper*.
- De Paola, M., & Scoppa, V. (2012). The effects of managerial turnover: evidence from coach dismissals in Italian soccer teams. *Journal of Sports Economics*, 13(2), 152–168.
- Gamson, W. A., & Scotch, N. A. (1964). Scapegoating in baseball. *American Journal of Sociology*, 70, 69–72.

- Gauriot, R., & Page, L. (2019). Fooled by performance randomness: overrewarding luck. *Review of Economics and Statistics*, 101(4), 658–666.
- Giambatista, R., Rowe, W., & Riaz, S. (2005). Nothing succeeds like succession: a critical review of leader succession literature since 1994. *The Leadership Quarterly*, 16(6), 963–991.
- Grusky, O. (1963). Managerial succession and organizational effectiveness. *American Journal of Sociology*, 69(1), 21–31.
- Heuer, A., Müller, C., Rubner, O., Hagemann, N., & Strauss, B. (2011). Usefulness of dismissing and changing the coach in professional soccer. *PloS one*, 6(3).
- Hughes, M., Hughes, P., Mellahi, K., & Guermat, C. (2010). Short-term versus long-term impact of managers: evidence from the football industry. *British Journal of Management*, 21(2), 571–589.
- Huson, M. R., Malatesta, P. H., & Parrino, R. (2004). Managerial succession and firm performance. *Journal of Financial Economics*, 74(2), 237–275.
- Jenter, D., & Kanaan, F. (2015). CEO turnover and relative performance evaluation. *Journal of Finance*, 70(5), 2155–2184.
- Kaplan, S. N., & Minton, B. A. (2012). How has CEO turnover changed? *International Review of Finance*, 12(1), 57–87.
- Kattuman, P., Loch, C., & Kurchian, C. (2019). Management succession and success in a professional soccer team. *PloS one*, 14(3).
- Koning, R. (2003). An econometric evaluation of the effect of firing a coach on team performance. *Applied Economics*, 35(5), 555–564.
- Lefgren, L., Platt, B., & Price, J. (2015). Sticking with what (barely) worked: a test of outcome bias. *Management Science*, 61(5), 1121–1136.
- Madum, A. (2016). Managerial turnover and subsequent firm performance: evidence from Danish soccer teams. *International Journal of Sport Finance*, 11(1), 46–62.

- Muehlheusser, G., Schneemann, S., & Sliwka, D. (2016). The impact of managerial change on performance: The role of team heterogeneity. *Economic Inquiry*, 54(2), 1128–1149.
- Muehlheusser, G., Schneemann, S., Sliwka, D., & Wallmeier, N. (2018). The contribution of managers to organizational success: evidence from German soccer. *Journal of Sports Economics*, 19(6), 786–819.
- Peeters, T., Szymanski, S., & Terviö, M. (2017). The inefficient advantage of experience in the market for football managers. *Working Paper*.
- Pieper, J., Nüesch, S., & Franck, E. (2014). How performance expectations affect managerial replacement decisions. *Schmalenbach Business Review*, 66(1), 5–23.
- Rowe, W. G., Cannella Jr, A. A., Rankin, D., & Gorman, D. (2005). Leader succession and organizational performance: Integrating the common-sense, ritual scapegoating, and vicious-circle succession theories. *The Leadership Quarterly*, 16(2), 197–219.
- Ter Weel, B. (2011). Does manager turnover improve firm performance? Evidence from Dutch soccer, 1986–2004. *De Economist*, 159(3), 279–303.
- van Ours, J. C., & van Tuijl, M. A. (2016). In-season head-coach dismissals and the performance of professional football teams. *Economic Inquiry*, 54(1), 591–604.